APPEARED IN BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 26, Number 1, Jan 1992, Pages 160-171

BOOK REVIEW

Large deviation techniques in decision, simulation, and estimation, by James A. Bucklew. John Wiley & Sons, 1990, 270 pp., $49.95. ISBN 0-471-61856-X

1. Deficiencies and virtues

In more ways than one, Professor Bucklew's book is a large deviation among all probability texts I have ever seen. The book has two goals: to present the basic results of a rapidly expanding branch of probability known as the theory of large deviations and to apply the theory to a number of problems arising in electrical engineering. Unlike all other texts on large deviations that have been published to date, this one is addressed to the engineering community, not to mathematicians. With this audience in mind, the author remarks in the preface that the theory of large deviations has a great potential for applications but is "unusually technical." His intention is to expose this material mostly by means of heuristic arguments while at the same time presenting "the main ideas and possible groundwork for the full-blown treatment." Approximately, half the book is devoted to applications. Given the audience to whom the book is addressed, it certainly sounds like an excellent plan.

Professor Bucklew has set himself a formidable task. It is one thing to write a technically correct book that has precise statements and complete proofs of all the results. Most books in mathematics are of this type. It is another thing—perhaps even more demanding—to write a book that omits complete proofs but motivates all the results by clear heuristic arguments.[1] Composing successfully such a text requires that the author have such a deep insight into the mathematics that he or she knows what to emphasize and what to omit. Such a text would be a very useful addition to the large deviation literature. I am afraid, however, that Professor Bucklew has not written it.

This book is seriously flawed. In many places, the expository and technical styles of presentation can only be described as sloppy. The level of mathematical rigor is much too low. Although essentially all of the mathematical results used in the book are readily available in the literature, too many of these results are misquoted; applying them as stated could lead to wrong answers. Key ideas in the theory are not given due emphasis. Important connections among various facets of the theory are not made. Errors, inconsistencies, and misleading statements abound. I will substantiate these harsh judgments by means of examples in the course of the

[1] Mark Kac was one of the masters of this type of exposition.

review. My opinion is that notwithstanding the book's wealth of fascinating topics and applications, it should not have been published in its present form.

The making of a book, even a flawed one, is a tedious task. So I write this review with a heavy heart. In fact, I would have preferred not to have written it at all except for the obligation I feel to the scientific community. The review is dedicated to all newcomers to the theory of large deviations who after reading it will be forewarned about using this book as a learning tool.

I want to be fair by pointing out the book's virtues up front. Although in my opinion the book is not suitable as an introduction to the theory, it is probably fine for experts in the theory of large deviations seeking new and/or interesting applications, in which the book abounds. Here is a sample: autoregressive processes (pp. 19–20), a theorem of H. Chernoff concerning statistical hypothesis testing (pp. 32–35), detection theory (Chapter VI), quick simulation (Chapter VIII), parameter estimation (Chapter IX), information theory (Chapter X). Chapters II-X contain sixty exercises that are all worked out in some detail. Seven appendices contain useful background material on analysis, probability, and statistics. Many of the mathematical derivations have a unifying theme, which is a large deviation theorem for random vectors first proved by J. Gärtner and later extended in [2]. The book presents useful heuristic derivations—containing, however, a number of inaccuracies—of several important large deviation results for stochastic processes in the literature, including Gaussian processes, dynamical systems perturbed by white noise, and slow Markov walks. Finally, the Notes and Comments at the end of each chapter provide useful information.

The reader may gain a hint about the problems contained in the text no later than in the third paragraph of the preface. The author states he is "content that the 'handwaving' sketches of some results contain the essential ideas and might be made rigorous with sufficient amounts of real analysis and measure theory." A key word in this statement is *might*, which suggests to me uncertainty on the part of the author. Already, the reader is left wondering whether the sketches can be made rigorous or not.

This doubt is not dispelled by reading Chapters II–V, which are the mathematical core of the book. The numerous glitches contained here caused me to doubt the accuracy of those theorems and applications in the book with which I am less familiar. The existence of so many errors, inconsistencies, and misleading statements in a book of this type is obviously a serious hindrance for the newcomer who might be unable to locate or correct them. At this point, a counterargument can be made. Since according to the preface the main audience is students of information theory, communication systems, applied statistics, and statistical signal processing, why should I make such a fuss about mathematical rigor and clarity of exposition? My opinion is that a textbook "even" for engineering students should make every attempt to be meticulous—though not necessarily rigorous—in the presentation of mathematical ideas. After all, the goal of this textbook is to present a new area of probability to people who will apply the theory to real world problems. Carrying out such applications inevitably entails making approximations to the existing theory. This is a hopeless task if the existing theory is not meticulously presented.

A specific instance of this kind of problem is Example 2 in Chapter IV, which treats a stochastic process with a boundary. The author states incorrectly that "these 'boundary' problems are of a technical nature and can be taken care of by successive approximation techniques." After he makes this statement, he indicates that he wants to study a specific event in which the boundary can essentially be ignored. He then proceeds to apply a large deviation theorem of A. D. Wentzell that is valid only for processes *without* boundaries but that gives the correct answers for processes with boundaries when applied to events in which the boundary is not involved. It is all terribly confusing. I am afraid that a newcomer to the theory of large deviations will get the impression that boundaries can be disregarded altogether, regardless of the events under consideration. This, of course, is completely false. In addition, the large deviation theorem of Wentzell that is applied in this example is itself misstated in a number of serious ways. Details will be given in Subsection 3c of this review.

Before proceeding with the discussion of the book, I will summarize some key ideas in the theory of large deviations.

2. Some key ideas in the theory of large deviations

The theory of large deviations studies the exponential decay of probabilities associated with certain random systems. An elementary example is provided by the sample means of independent, identically distributed (i.i.d.) random variables $\{X_i, i \in \mathbf{N}\}$ that are exponentially bounded. The latter holds if for all $\alpha \in \mathbf{R}$

$$E\{\exp(\alpha X_i)\} < \infty,$$

where $E\{\cdot\}$ denotes expected value. Let m equal the expected value $E\{X_1\}$ of the random variable X_1 and define the sample means

$$S_n/n = \sum_{i=1}^{n} X_i/n \quad \text{for } n \in \mathbf{N}\,.$$

The weak law of large numbers states that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|S_n/n - m| \geq \varepsilon\} = 0\,. \tag{2.1}$$

In 1938, H. Cramér showed that the rate of decay of the probabilities in (2.1) is in fact exponentially fast and gave a formula for the exponential decay rate. This result, the first in the theory of large deviations, has been extensively generalized to cover many new situations. These situations include i.i.d. random vectors taking values in $\mathbf{R}^d$ and in certain infinite dimensional spaces; Markov chains and Markov processes; Gaussian processes; stochastic processes arising in numerous areas of application including statistical mechanics, interacting particle systems, statistics, and information theory.

A general framework is needed to encompass these many situations. Let $\mathcal{X}$ be a complete separable metric space and $\{Y_n, n \in \mathbf{N}\}$ a sequence of random vectors taking values in $\mathcal{X}$. Typically, the sequence Y_n converges in probability to a point x_0 in $\mathcal{X}$ as $n \to \infty$. Thus, for any Borel subset A of $\mathcal{X}$ whose closure does not contain x_0, the sequence of probabilities

$$Q_n\{A\} \doteq P\{Y_n \in A\} \tag{2.2}$$

converges to 0 as $n \to \infty$. The theory of large deviations addresses the natural question whether these probabilities converge to 0 exponentially fast and if so seeks to express the decay rate as a function of the set A. One of the pleasant surprises of the theory is that the resulting expression often involves quantities that are of fundamental importance for the underlying problem.

The exponential decay of the probabilities in (2.2) is abstracted in the key concept of large deviation principle. Let $\{Q_n, n \in \mathbf{N}\}$ be a sequence of probability measures on the Borel subsets of $\mathcal{X}$. Let I be a function mapping $\mathcal{X}$ into $[0, +\infty]$. For any subset A of $\mathcal{X}$, we write $I(A)$ for the quantity $\inf_{\beta \in A} I(\beta)$. The sequence $\{Q_n\}$ is said to satisfy the *large deviation principle* (or LDP) *with rate function I* if the following three conditions hold.

(a) For each $s \in \mathbf{R}$ the level set

$$\Phi(s) \doteq \{\beta \in \mathcal{X} \colon I(\beta) \leq s\} \tag{2.3}$$

is compact.

(b) For each closed set C in $\mathcal{X}$, the *upper large deviation bound* holds:

$$\limsup_{n \to \infty} n^{-1} \log Q_n\{C\} \leq -I(C)\,. \tag{2.4}$$

(c) For each open set G in $\mathcal{X}$, the *lower large deviation bound* holds:

$$\liminf_{n \to \infty} n^{-1} \log Q_n\{G\} \geq -I(G)\,. \tag{2.5}$$

Cramér's theorem proves that the distributions of the sample means $\{S_n/n\}$ of i.i.d. exponentially bounded random variables satisfies the LDP. The rate function is given by the formula

$$I(\beta) = \sup_{\alpha \in \mathbf{R}} \{\alpha\beta - h(\alpha)\}, \tag{2.6}$$

where $h(\alpha), \alpha \in \mathbf{R}$, denotes the finite convex function

$$h(\alpha) = \log E\{\exp(\alpha X_1)\}.$$

Formula (2.6) exhibits $I(\beta)$ as the Legendre-Fenchel transform of the convex function $h(\alpha)$. Thus, $I(\beta)$ is convex and $h(\alpha)$ may be recovered from $I(\beta)$ by reiterating the transform; viz., $h(\alpha) = \sup_{\beta \in \mathbf{R}} \{\alpha\beta - I(\beta)\}$.

In 1966, S. R. S. Varadhan discovered an important application of the LDP to the asymptotic evaluation of certain integrals. Let F be a bounded continuous function mapping $\mathcal{X}$ into $\mathbf{R}$. He proved that if the sequence $\{Q_n\}$ satisfies the LDP with rate function I, then

$$\lim_{n \to \infty} n^{-1} \log \int_{\mathcal{X}} \exp[nF(\beta)] Q_n(d\beta) = \sup_{\beta \in \mathcal{X}} \{F(\beta) - I(\beta)\}\,. \tag{2.7}$$

This fact is easily motivated. If we summarize the upper and lower large deviation bounds (2.4) and (2.5) by the heuristic formula

$$Q_n(d\beta) \asymp \exp[-nI(\beta)]\, d\beta,$$

then the limit in (2.7) should equal

$$\lim_{n\to\infty} n^{-1} \log \int_{\mathcal{X}} \exp(n[F(\beta) - I(\beta)])\, d\beta.$$

By analogy with Laplace's method, the latter limit should equal

$$\sup_{\beta\in\mathcal{X}} \{F(\beta) - I(\beta)\}.$$

For the purpose of reference later in this review, I will comment on each of the three conditions in the definition of LDP. Condition (a) is extremely useful in many situations. It implies that $I(\beta)$ is lower semicontinuous and guarantees that a rate function is unique. It also assures that $I(\beta)$ attains its infimum on any closed set in the space. Conditions (b) and (c) applied to the closed-open set $\mathcal{X}$ show that the infimum of $I(\beta)$ over the whole space $\mathcal{X}$ equals 0. Because of condition (a), we know that this infimum is attained on a nonempty set of points—call it $\mathcal{E}$—and that if A is any Borel subset of $\mathcal{X}$ such that the closure of A is disjoint from $\mathcal{E}$, then the quantity $I(A)$ is positive. By applying the upper large deviation bound (2.4) to the closure of A, we conclude that the probabilities $Q_n\{A\}$ converge to 0 exponentially fast. The set $\mathcal{E}$ often plays an important role in applications. For example, in the study of statistical mechanical models, $\mathcal{E}$ may be associated with the set of equilibrium states of the model. This list of consequences of condition (a)—which may be augmented by examining any rigorous text on large deviations—should highlight its importance.

A natural question closely related to the upper and lower large deviation bounds (2.4) and (2.5) is to find conditions on a Borel set A in $\mathcal{X}$ such that the following *large deviation limit* exists:

$$\lim_{n\to\infty} n^{-1} \log Q_n\{A\} = -I(A)\,. \tag{2.8}$$

We denote by $\overline{A}$ and by A° the closure of A and the interior of A, respectively. Since $\overline{A} \supset A \supset A^\circ$, the upper and lower large deviation bounds applied to the respective sets $\overline{A}$ and A° show that

$$-I(A^\circ) \le \liminf_{n\to\infty} n^{-1} \log Q_n\{A\} \le \limsup_{n\to\infty} n^{-1} \log Q_n\{A\} \le -I(\overline{A})\,.$$

We conclude that the large deviation limit (2.8) holds provided

$$I(A^\circ) = I(\overline{A})\,. \tag{2.9}$$

Any Borel set A satisfying this condition is called an *I-continuity set.*

The key large deviation theorem that is applied numerous times in the book under review is the so-called Gärtner-Ellis theorem, which generalizes Cramér's theorem. Let $\{Y_n, n \in \mathbf{N}\}$ be a sequence of random vectors taking values in $\mathbf{R}^d$ and for $\alpha \in \mathbf{R}^d$ consider the limit

$$h(\alpha) = \lim_{n\to\infty} n^{-1} \log E\{\exp\langle\alpha, nY_n\rangle\},$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbf{R}^d$. Under the assumption that $h(\alpha)$ exists, is finite for all $\alpha \in \mathbf{R}^d$, and is differentiable on $\mathbf{R}^d$, the paper of J. Gärtner [5] proves that the distributions of $\{Y_n\}$ satisfy the LDP with rate function $I(\beta) = \sup_{\alpha \in \mathbf{R}^d}\{\langle \alpha, \beta \rangle - h(\alpha)\}$. Gärtner's result was extended in [2] to cover the case where $h(\alpha)$ is finite only on a proper subset of $\mathbf{R}^d$. If Y_n equals the nth sample mean of i.i.d. exponentially bounded random variables, then the Gärtner-Ellis theorem reduces to Cramér's theorem. A slightly stronger form of the theorem was found independently by M. I. Freidlin and A. D. Wentzell and is proved in §5.1 of [4].

With these key ideas out of the way, it is time to return to the book under review.

3. Detailed comments on chapters II–V

The introductory first chapter of the book discusses for an engineering audience the applicability of large deviation ideas to problems arising in the study of communication systems. I will comment in detail on the next four chapters, II–V, which are the mathematical core of the book. The remainder of the text is devoted to applications and further developments of the theory: VI. *Applications to detection theory*, VII. *Asymptotic expansions*, VIII. *Quick simulation*, IX. *Applications to parameter estimation*, X. *Applications to information theory.* It is conceivable that these last five chapters are written with greater care than the ones I read closely. But I must admit that having spent considerable effort trying to straighten out Chapters II–V, I ran out of steam.

This section of the review serves two purposes. First, my claim that this book is seriously flawed will be substantiated by citations of specific problems. Second, I will try in many cases to indicate how these problems may be straightened out—here, I am thinking of the newcomer to the theory. The reader with book in hand may easily follow the details. Anyone without the book will easily be able to get the flavor just by reading on.

3A. Chapter II. Cramér's Theorem and extensions

This chapter sketches proofs of H. Cramér's large deviation theorem of 1938 and the extensions of the theorem carried out first by J. Gärtner and later in [2]. The presentation is seriously marred by the author's failure to define fully and precisely the concept of large deviation principle (LDP). In each of the large deviation theorems presented, the author gives the upper and lower large deviation bounds but neglects to mention the standard condition that the rate function have compact level sets. The latter is condition (a) in the definition of LDP, as discussed in §2 of this review. Condition (a) and the condition that the rate function be lower semicontinuous (condition (a) implies the second) are dismissed in the Notes and Comments as being "technical considerations [that] will rarely, if every [sic], be necessary." Unfortunately, the author proves himself wrong. To have included condition (a), which should not be considered technical at all, would have caused much less difficulty than to have avoided it.

There are at least three major instances in which mishandling condition (a) causes trouble. The first involves the numerous consequences of condition (a) noted in §2 of this review, many of which are very useful in applications. For example, without condition (a), one cannot conclude that if A is a closed set such that

$I(x) > 0$ for all $x \in A$, then the probabilities $Q_n\{A\}$ converge to 0 exponentially fast.

The second appears in Chapter IV. In two separate theorems stated on pp. 61 and 68, the large deviation bounds are stated, not for arbitrary closed and open subsets of the associated metric space, but in the terms of a formulation due to M. I. Freidlin and A. D. Wentzell and presented in §3.3 of [4]. These authors give a lower large deviation bound for arbitrary open balls, which is clearly equivalent to the lower large deviation bound (2.5). They also give an upper large deviation bound for certain closed sets defined in terms of the level sets $\{\Phi(s)\}$ of the rate function. Under condition (a), this upper bound is equivalent to the upper large deviation bound (2.4). Without it, the upper bounds in the theorems on pp. 61 and 68 do not imply the upper large deviation bounds (2.4). Thus, as the theorems are stated, it does not follow that the rate functions are the given functions $I(f)$.[2]

The third instance related to condition (a) arises in Chapter V, in the statement and proof of Varadhan's theorem on the asymptotic evaluation of certain integrals (pp. 84–85). This theorem was mentioned in §2 of this review. Here the book makes a misleading statement in the opposite direction. Although for the first time in the text condition (a) is included, unfortunately it is not needed in the proof of Varadhan's theorem at all. The boundedness and continuity of the function F together with the upper and lower large deviation bounds suffice.

Besides neglecting to give the full definition of large deviation principle, Chapter II makes another important omission by neglecting to isolate the concept of an I-continuity set. We recall that if the set satisfies the continuity condition (2.9), then the large deviation limit (2.8) holds. Because of this omission, there are a number of places where the author would like to deduce the existence of a large deviation limit, but is unable to. These places are either marked with the awkward parenthetical phrase "assuming limits exist"—see pp. 33, 51, 66 and 112—or are not indicated at all—see pp. 55, 62, and 93.

On p. 13 the author is wrong to imply that any convex set in $\mathbf{R}^d$, $d \in \mathbf{N}$, is an I-continuity set. Indeed, condition (2.9) is satisfied for all convex sets having nonempty interior if and only if the rate function is finite on the whole space; in general, this does not hold.

There are some other minor problems in Chapter II. On p. 7, the moment generating function $M(\theta)$ must be assumed to exist. Contrary to the statement on p. 8, a convex function I on $\mathbf{R}$ is not in general continuous everywhere on the set $\{x \in \mathbf{R}\colon I(x) < \infty\}$, but only on the interior of this set.[3] The general relationship between the set $\{x \in \mathbf{R}\colon I(x) < \infty\}$ and the support of the random variable x_1 is not mentioned in the chapter[4] although this information would have been useful in understanding the assumptions in the book's somewhat unorthodox formulation of the Gärtner-Ellis theorem on pp. 15–16. If this general relationship had been mentioned, then the reader would have been alerted to the fact that the rate function in Cramér's theorem takes the value $+\infty$ on some nonempty interval whenever the random variable x_1 is bounded above or below. Contrary to the statement on p. 9, the finiteness of the moment generating function on all of $\mathbf{R}$ requires more than just the finiteness of moments of all orders.

[2] That these functions have compact level sets is shown in the papers of A. D. Wentzell [6, 7].

[3] The correct statement is on p. 183 of Appendix B.

[4] This is well known. For example, see Theorem VIII.3.1 in [3]. A special case of this relationship is treated in Property 3 on p. 8 of the book under review.

While I did not examine most of the exercises closely, I found errors in two that I looked at. The random variables $\mathbf{Y}_n$ in Exercises 4 and 5 on p. 24 are ambiguously defined and the solution of Exercise 5 is incorrect; indeed, $\phi(\theta) = |\theta| - \log 2$ for *all* $\theta \in \mathbf{R}$.

3B. Chapter III. Sanov's Theorem and the contraction principle

This chapter develops the large deviation theory—first investigated by I. Sanov—of the empirical distributions of i.i.d. random vectors taking values in $\mathbf{R}^d$, $d \in \mathbf{N}$. The empirical distributions take values in the set of probability measures on $\mathbf{R}^d$, which is the space on which the associated large deviation phenomena must be studied. The rate function is the important quantity known as the relative entropy or the Kullback-Leibler information number. The chapter ends with a discussion of the contraction principle, which is a useful technique for deriving a new LDP from a known LDP when the underlying probability measures are related by a continuous mapping.

Although this chapter is cleaner than Chapter II, it does contain a number of glitches. It would have helped the reader if the author had explained why the maximizing θ on pp. 27 can be found by differentiation (reason: convexity). Almost half of p. 28 is devoted to showing that in the case of i.i.d. random vectors taking values in a finite set of cardinality d, the rate function $I(x)$ for the corresponding empirical distributions equals $+\infty$ whenever x is not a probability vector in $\mathbf{R}^d$. However, this follows immediately and with hardly any calculation from the lower large deviation bound applied to the open set that is the complement in $\mathbf{R}^d$ of the set of probability vectors.[5] The rate function in the theorems on pp. 28 and 36 is not identified by name—the Kullback-Leibler information number—until p. 92 in Chapter VI.[6] The second display on p. 33 does not seem to fit. Despite the statement at the end of the first paragraph on p. 37, no reference is given in the Notes and Comments concerning the LDP for the quantities $\{\mathbf{L}_n^k\}$. The book does not mention an important application of the contraction principle on p. 38, which is the relationship between the rate function for the sample means of i.i.d. bounded random variables and the rate function for the corresponding empirical distributions. This would have been natural because of the discussion on p. 29 relating these two random quantities.

3C. Chapter IV. Gaussian processes and Wentzell-Freidlin theory

This chapter presents heuristic derivations of large deviation theorems for a number of important stochastic processes appearing in the literature: Gaussian processes, dynamical systems perturbed by white noise, and slow Markov walks in discrete time and in continuous time. This chapter also works out a number of examples that are intended to illustrate how the theorems may be applied.

I will start with what I consider to be the most poorly written portion of the book: the theorem of A. D. Wentzell stated on p. 61 and Example 2 on pp. 63–65, in which the theorem is supposedly applied. The theorem is a large deviation result concerning slow Markov walks in discrete time.

[5] For example, see p. 252 of [3].

[6] The term "relative entropy" appears in the Notes and Comments at the end of Chapter III, but it is not made clear to what this phrase refers. On p. 160, the rate function is identified again but the word "information" is dropped.

Comments concerning Wentzell's theorem on p. 61. The presentation has four problems:

(a) The theorem is preceded by four extremely technical conditions that correspond respectively to Conditions C, D, E, and F in the paper of Wentzell [7]. However, Condition 1 in the book lops off the last two hypotheses in Wentzell's Condition C, including the most serious and complex restriction of all, which is a continuity condition on the statistics of the underlying Markov chain.[7] It is precisely this condition that is violated by stochastic processes with boundaries. In particular, it is violated in Example 2, in which the theorem is supposed to be applied.

(b) In the statement of the theorem, the symbols $\limsup_{n\to\infty}$ and $\liminf_{n\to\infty}$ must be interchanged. The former must go with the upper bound and the latter with the lower bound.

(c) The large deviation bounds in this theorem are given in the terms of the Freidlin-Wentzell formulation. Since the book neglects to include the condition that the rate function have compact level sets, the upper bound in this theorem does not imply the upper large deviation bound (2.4). Thus, as the theorem is stated, it does not follow that the rate function is the given function $I(f)$ (see footnote 2).

(d) The book makes no attempt to motivate the four extremely technical conditions that precede the statement of the theorem. In addition, although this is the first appearance in the book of Freidlin-Wentzell-type bounds, no comments are made concerning them, even just to warn the reader that the large deviation bounds are not being given in the form used in the first sixty pages.

Comments concerning Example 2 on pp. 63–65. This example involves a discrete Markov chain on the nonnegative integers with a boundary at 0. Before pointing out what feature of the process he is interested in, the author states that because of the boundary behavior, "the process will violate some of the conditions of the theorem of Wentzell." A parenthetical comment states that a specific function is not differentiable. True enough, but this is not one of the four conditions as given in the text. In fact, as pointed out above, the violated condition is one that the author erroneously omitted. This is followed by the incorrect statement—already noted in §1 of this review but worth repeating—that "these 'boundary' problems are of a technical nature and can be taken care of by successive approximation techniques." The author then points out that he is interested in the probability of the event that the Markov chain, starting from 0, reaches a high level N before hitting 0 again. Rescaling the process, the author calculates the asymptotic probability of the rescaled event as $N \to \infty$ by applying the theorem of Wentzell, which is intended for processes without boundaries. He is saved—and obtains the correct answer for the special case considered in Example 2—because for this particular event, an approximation argument based on a random walk without boundary (to which Wentzell's theorem applies) renders the effect of the boundary negligible in the limit $N \to \infty$. However, this is not at all spelled out clearly.

Another way to have avoided the confusion in Example 2 would have been to note that the scaled process in this example satisfies the LDP but with a rate function different from the one given in the book (see Theorem 5.1 in [1]).

[7] Professor Wentzell devotes an entire section of his paper [7] to the discussion of this condition.

Chapter IV has a number of other problems. The * notation on pp. 44–46 to denote l_2 counterparts to $L_2[0,T]$ quantities is confusing because of its standard use to denote dual spaces and adjoint operators. The quantities $\{\mathbf{a}_k\}$ on line 5 of p. 45 should be identified as *zero mean* independent Gaussian random variables. An "inverse kernel" is defined in the third display on p. 46 and appears in the rate function in the theorem attributed by the author to M. Schilder. Without the condition that the covariance function be positive definite, which is equivalent to the condition that all the eigenvalues $\{\lambda_k\}$ be positive, the definition of the inverse kernel makes no sense. However, the author neglects to state this condition. The theorem on p. 46 attributed to M. Schilder was in fact first proved by M. Schilder for Brownian motion and was later extended by M. Pincus to Gaussian processes having continuous paths and having covariance functions that are continuous and positive definite. Unfortunately, none of these hypotheses appears in the book's statement of the theorem. This misquoting of the theorem foreshadows the even more serious misquoting of Wentzell's theorem fifteen pages later.

A theorem of M. I. Freidlin and A. D. Wentzell on the asymptotics of the mean exit time of certain dynamical systems perturbed by white noise is stated on p. 54. The author does not discuss the significance of the second hypothesis involving the exterior normal. I had great difficulty following the heuristic explanation of this theorem given in the text.

In the motivation that precedes Wentzell's theorem on p. 61, the author refers to the "jump sizes" of the process $\mathbf{x}^n(t)$ (p. 56). However, the latter process is defined by linear interpolation and has continuous paths; it has no jumps at all.

Examples 3 and 4 on pp. 70–72 are slightly more careful than Example 2. A large deviation theorem that holds only for processes without boundaries is applied to processes with boundaries in order to study events in which the effect of the boundary can be rendered negligible by an approximation argument. While the argument is not spelled out, at least in these examples the author recognizes the problem that the boundaries pose. By reversing the order of the presentation from that in Example 2, the author does not imply that the theorem may be used to study all large deviation phenomena involving these processes with boundaries, but only that the conclusions of the theorem can be made applicable in the particular situations considered.

The large deviation theorem that is supposedly applied in Examples 3 and 4 is stated on p. 68. As in the erroneous statement of Wentzell's theorem on p. 61, the symbols $\limsup_{n\to\infty}$ and $\liminf_{n\to\infty}$ should also be interchanged here. There is also an issue of consistency. The symbol $\limsup_{\varepsilon\to 0^+}$ in the lower bound (which should be $\liminf_{\varepsilon\to 0^+}$) and the symbol $\liminf_{\varepsilon\to 0^+}$ in the upper bound (which should be $\limsup_{\varepsilon\to 0^+}$) are unnecessary. As on p. 61, it suffices if the bounds hold for all $\varepsilon > 0$.

3D. Chapter V. Large deviations for Markov processes

This chapter treats the large deviation theory of three related processes: (a) the sample means of functionals of finite state Markov chains in discrete time, (b) the sample means of functionals of finite state Markov chains in continuous time, and (c) the empirical distributions of finite state Markov chains in discrete time.

Errors, inaccuracies, and misleading statements continue to appear. The Perron-Frobenius theorem and the Gärtner-Ellis theorem are used to treat case (a), in which the Markov chain must be aperiodic and irreducible. These conditions guarantee that the operator T_θ defined at the bottom of p. 76 is primitive—i.e., that some power of T_θ is positive—and that the key asymptotic relation given near the bottom of p. 77 holds.[8] Inexplicably, the author omits the condition of aperiodicity both in the initial discussion on p. 76 and in the statement of the theorem on p. 78 although he includes it in Appendix C.4, where the Perron-Frobenius theorem is summarized (but without any comment concerning the asymptotic relation used on p. 77). Without giving any reason why, the author states on p. 78 the nontrivial fact that the function $\log \lambda(\theta)$ satisfies the steepness condition of the Gärtner-Ellis theorem.

Case (b) is treated heuristically. Although on p. 80 it is stated that the form of the rate function will be determined, no formula for the rate function is given. It is left to the reader to figure out that one must substitute $\lambda(\theta)$ in his formula (15) back into his formula (7), which appears under case (a).

Concerning case (c), the author neglects to give a full proof of the large deviation principle. The latter may be derived as in case (a).[9] Also as in case (a), the condition of aperiodicity of the Markov chain is not mentioned along with the stated condition of irreducibility. Finally, the author fails to point out that the LDP for case (a) may be easily derived from the LDP for case (c) via the contraction principle. This would give a second variational formula for the rate function in the former case besides formula (7) in the book.

4. Summary

It is a pity that Professor Bucklew did not eliminate the numerous errors, inconsistencies, and misleading statements in his book. If he had, he would have produced not this highly flawed text, but a valuable contribution to the large deviation and engineering literature.

Acknowledgment

I am very grateful to Paul Dupuis for help in preparing this review.

References

1. P. Dupuis, *Large deviations analysis of some recursive algorithms with state dependent noise*, Ann. Probab. **16** (1988), 1509–1536.
2. R. S. Ellis, *Large deviations for a general class of random vectors*, Ann. Probab. **12** (1984), 1–12.
3. R. S. Ellis, *Entropy, large deviations and statistical mechanics*, Springer-Verlag, New York, 1985.
4. M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, Springer-Verlag, New York, 1984.
5. J. Gärtner, *On large deviations from the invariant measure*, Theory Probab. Appl. **22** (1977), 24–39.
6. A. D. Wentzell, *Rough limit theorems on large deviations for Markov stochastic processes*, I, Theory Probab. Appl. **21** (1976), 227–242.

[8] This relation is not stated precisely, but in terms of the symbol $\approx$, which is used repeatedly in the book to denote a heuristic derivation.

[9] For example, see §III of [2].

7. ———, *Rough limit theorems on large deviations for Markov stochastic processes*, II, Theory Probab. Appl. **21** (1976), 499–512.

Richard S. Ellis
University of Massachusetts